Simon Müller, Christian Korff, Dietrich Manzey

Moving-horizon versus moving-aircraft: Effectiveness of competing attitude indicator formats on recoveries from discrete and continuous attitude changes

Subtitle

Müller, S., Korff, C., & Manzey, D. (2020). Moving-horizon versus moving-aircraft: Effectiveness of competing attitude indicator formats on recoveries from discrete and continuous attitude changes. Journal of Experimental Psychology: Applied. https://doi.org/10.1037/xap0000329

Moving-Horizon Versus Moving-Aircraft: Effectiveness of Competing Attitude Indicator Formats on Recoveries From Discrete and Continuous Attitude Changes

Simon Müller, Christian Korff, and Dietrich Manzey

Chair Work, Engineering & Organizational Psychology,

Department of Psychology and Ergonomics, Technische Universität Berlin

Author Note

Many thanks to all experimental participants and Stephan Pietschmann for his software support.

Correspondence concerning this article should be addressed to Simon Müller, Chair Work, Engineering & Organizational Psychology, Department of Psychology and Ergonomics, Technische Universität Berlin, Marchstr. 12, 10587 Berlin, Germany.

Email: simon.mueller@tu-berlin.de

Abstract

The present research revisits the old issue whether attitude information is best conveyed to pilots in a *moving-horizon* format or in a *moving-aircraft* format. Previous research has suggested that the moving-aircraft format might not be beneficial for flight path tracking but recoveries from unusual attitudes, although the result are not fully consistent. A limitation of studies to date is that the recovery task usually did not involve progressive attitude changes of the aircraft but only sudden discrete changes. Compared with a discrete stimulus, the continuous dynamics might increase the perceived time pressure to respond, which in turn can be expected to amplify the error proneness with a less intuitive format. Two experiments were conducted where flight novices and experienced pilots performed tracking and recover tasks with both formats. Recoveries were performed from both, sudden (discrete) and continuously developing attitude changes. Independent of whether novices or pilots were considered, the general superiority of the moving-aircraft format was confirmed. As expected, the benefits of this format became even more evident with progressive attitude changes. No differences were found for tracking. The results add to the evidence favoring the moving-aircraft over the moving-horizon format for both novices and pilots. The moving-aircraft format of the attitude indicator should at least be considered as a standard for new applications, such as ground control stations of unmanned aerial vehicles.

Public Significance Statement: The present studies confirm the general superiority of the moving-aircraft over the moving-horizon reference format of attitude indicators in aircraft cockpits. In addition, it shows that this superiority is even more pronounced for pilots and novices when comparing recovering from realistic continuous attitude changes compared with commonly used sudden discrete changes. The results add new arguments to the case that an adaption of moving-aircraft displays as new standard for aircraft cockpits would be reasonable and should at least be considered for ground control stations of unmanned aerial vehicles.

Keywords: flight displays, display-control or stimulus-response compatibility, display design principles, expert-novice differences, unusual attitude recovery

Moving-Horizon Versus Moving-Aircraft: Effectiveness of Competing Attitude Indicator Formats on Recoveries From Discrete and Continuous Attitude Changes

Maintaining a proper mental representation of an aircraft's position and movement relative to the Earth's surface is of paramount importance for pilots (Previc & Ercoline, 2004). It requests a pilot's constant awareness of the aircraft's position relative to the natural horizon's position. To ensure the pilot's proper spatial orientation also in instrument meteorological conditions (i.e., when a natural horizon for visual reference is missing, e.g., flying above clouds or in low visibility), aircraft are equipped with an attitude indicator (AI). Two different formats of this AI can be distinguished, which have been referred to as *moving-horizon* (MH, aka *inside-out*) or *moving-aircraft* (MA, aka *outside-in*) format (see Figure 1). Whereas the former represents today's standard format used in commercial, civil, and military aviation, the latter has mainly been used by Russian aircraft manufacturers.[1]

The MH format shows an artificial horizon that is moving according to the perceived movements of the natural horizon when looking outside the cockpit windscreen in case of bank or pitch changes. By reference to this rolling movement, the bank angle of the aircraft is visualized. Banking of the aircraft to the left is indicated by turning the artificial horizon to the right and vice versa. The aircraft's pitch movements are depicted as up or downward movements of the horizon line; that is, a pitch-up corresponds to a downward movement of the horizon and vice versa. In the center of the display, a fixed

[1] Note that also other AI formats have been proposed, like the frequency-separated format (Beringer et al., 1975) or the Malcolm horizon (Malcolm, 1983). However, the MH and MA format are the only ones, which actually have been implemented for routine operations in commercial and/or military aircrafts.

airplane symbol is displayed as a stable reference. This format has always been the standard format used in all aircrafts manufactured in Europe and North America. The guiding (compatibility) principle underlying this design is the so-called *principle of pictorial realism* (Roscoe, 1968), stating that a cockpit display "should 'look like' or be a pictorial representation of the information that it represents" (Wickens, 2003, p. 152).

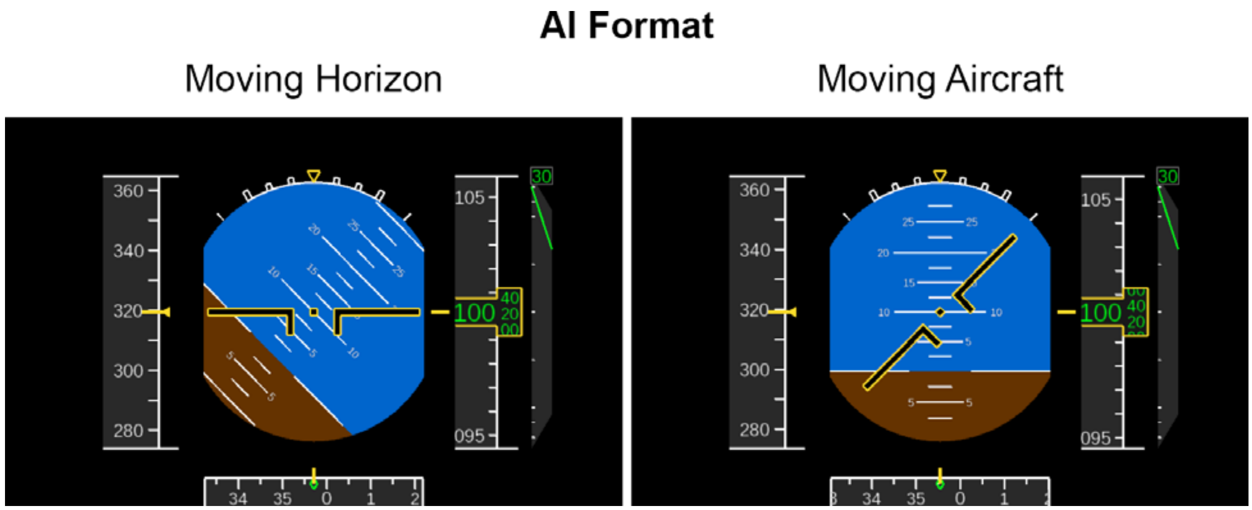

Figure 1. Illustration of the two attitude indicator (AI) reference formats. The left side shows the moving-horizon (MH) format and the right side the moving-aircraft (MA) format. Both primary flight displays (PFDs) show a bank angle of 45° to the left and pitch up of 10°. The instruments shown were also used in the experiments. See the online article for the color version of this figure.

The MA format also shows an aircraft symbol in the center and an artificial horizon line separating ground from sky. As with the MH format, the pitch angle of the aircraft is represented by shifting the artificial horizon line upward or downward. However, in contrast to the MH format, the angle of bank is indicated by turning the aircraft symbol, while the artificial horizon line remains in a horizontal position. This means that bank movements of the aircraft to the right or to the left are depicted directly by turning the aircraft symbol in the same direction. This format is based on the so-called *principle of the moving part* (Roscoe, 1968), stating that “the moving element on a display should correspond with the element that moves in the pilot’s ‘mental model’, or mental representation of the aircraft, and should move in the same direction as that mental representation” (Wickens, 2003, p. 152).

The human factors issue of which of these formats is better suited to provide an intuitive and quick-to-interpret indication of an aircraft’s attitude has been a matter of concern for a long time. The majority of the studies comparing these formats have been conducted in the 1950–1970s. They mostly investigated the performance of flight novices or pilots in recovery tasks; that is, the participants were required to recover as quickly and accurately as possible from unusual flight attitude to a horizontal flight (e.g., Roscoe & Williges, 1975). When investigating flight novices, a clear superiority of the MA over the MH format, in terms of quicker response times and/or less reversal errors, that is, incidents where participants initially steered in the wrong direction, were found. Results of rather rare studies with experienced pilots were not as consistent as with novices, though, with some studies showing advantages of the MA compared with the MH format (e.g., Dunlap & Associates, 1955; cited in Previc & Ercoline, 1999), some studies showing the reverse (e.g., Beringer, Williges, & Roscoe, 1975), but the majority of studies showing no clear differences between both AI formats (e.g., Browne, 1954; Roscoe & Williges, 1975). The latter result suggested that switching from the MH format to the unfamiliar MA format would not lead to significant performance impairments, even in experienced pilots (over-)trained with the MH format. Overall, the results of this time period suggest that the MA format of the AI might be more intuitive than the MH format, although the latter continued to be the standard AI format in most aircraft until today (cf. reviews by Johnson & Roscoe, 1972; Previc & Ercoline, 1999).

A number of theories have been raised to explain the potential superiority of the MA format specifically for recovery tasks (Previc & Ercoline, 1999). The most promising ones relate to the higher *response-effect compatibility* of the MA compared with the MH format. The response-effect compatibility is reflected in the fact that inputs at the sidestick or yoke to the right or left lead to direct corresponding (thus, compatible) movements of the aircraft symbol in the MA display, but directly opposite (thus, incompatible) movements of the artificial horizon in the MH display (Janczyk, Pfister, Crognale, & Kunde, 2012; Janczyk, Yamaguchi, Proctor, & Pfister, 2015). It has been shown that this involves extra cognitive effort in building a proper mental representation of the aircraft's position even for experienced pilots (Kovalenko, 1991). This theory implicitly assumes that pilots intuitively tend to interpret the movements of the display as what they control by their steering input. This is in line with the so called neuropsychological reference model proposed by Previc and Ercoline (1999), which predicts that objects in close proximity to the pilot, like cockpit instruments, will always be perceived and processed as controllable and moving objects, one can interact with, while large-scale movements in far distances of the field of view, like the natural horizon, are usually interpreted as consequences of self-motion. Overall this would mean that the compatibility principle of the moving part entails much stronger effects and becomes dominant over the rivaling principle of pictorial realism.

The issue has received new attention nowadays, mainly related to new generations of cockpit design and issues of designing ground stations for control of unmanned aerial vehicles (e.g., Beringer & Ball, 2009; Cohen, Otakeno, Previc, & Ercoline, 2001;

Ding & Proctor, 2017; Gross & Manzey, 2014; Lee & Myung, 2013; Müller, Roche, & Manzey, 2019; Müller, Sadovitch, & Manzey, 2018). The result of these more recent studies are less consistent than the previous research with respect to the effectiveness of the two formats. Some of these studies were in line with the results of the earlier studies and reaffirmed the superiority of the MA format (Ding & Proctor, 2017; Lee & Myung, 2013; Müller et al., 2018). However, others could not replicate the advantages of the MA compared with the MH format (Cohen et al., 2001; Gross & Manzey, 2014; Yamaguchi & Proctor, 2010).

This discrepancy may be attributable to several factors, including the flight tasks considered and the overall layout of the display used. For example, Cohen et al. (2001) and Yamaguchi and Proctor (2010) did not use recovery tasks to compare the performance consequences of the two AI formats, but a simulation of a flight path tracking task. In these simulations, participants were required to maintain a horizontal flight attitude despite of continuous deviations, simulating turbulent atmospheric conditions. Thus, this task involved the response to continuously occurring deviations from a horizontal attitude, which, however, were relatively small, requesting only small inputs at the control device. In contrast, the recovery tasks usually used in other research simulates a possibly dangerous flight situation in which the aircraft changes its attitude because of environmental impact, a flight system failure, or pilot induced causes. The pilot is surprised by this unexpected, extreme flight attitude change, needs to establish as quickly as possible a proper spatial orientation by looking at the AI, and must initiate a rapid compensatory flight maneuver, to recover to a horizontal attitude (e.g., Roscoe & Williges, 1975). In flight trainings, this task is usually called upset recovery. The student closes the eyes and lowers the head, while the instructor maneuvers to the unusual attitude. In laboratory studies, the "undetected" change of the flight attitude is often skipped, and the unusual altitude is presented as a sudden change in the AI. It is obvious, that this latter task involves higher demands on spatial orientation as a tracking task, which might be the reason that performance in recovery tasks is more sensitive to differences between the different AI formats than tracking performance. Support for this has also been provided by a recent study of Müller et al. (2018) who found the MA format superior than the MH format for tracking tasks only for flight novices but not experienced pilots.

Another factor responsible for the inconsistent findings in more recent research might be assumed in the chosen general layout of the used AI displays. While studies, which found benefits of the MA compared with the MH display, usually used displays resembling or even directly corresponding to typical primary flight displays (PFDs) actually implemented in today's aircrafts (Lee & Myung, 2013; Müller et al., 2018), studies not finding differences often used more artificial laboratory displays (Gross & Manzey, 2014; Yamaguchi & Proctor, 2010).

The present study again addresses the issue of what AI format is more intuitive and better suited to support spatial orientation and attitude control. Capitalizing on our previous research (Müller et al., 2018), we involved flight novices and experienced pilots who were requested to perform tracking and recovery tasks with support of different AI formats, implemented in primary flight displays, closely resembling those used in current generations of aircraft. However, going beyond our previous approach, we did not only use the conventional recovery task, which has been used in all previous research, that is, recovery from suddenly occurring discrete attitude changes, but also a task version where the attitude changes occurred dynamically (continuously) with different angular velocities. The rationale behind this variation was twofold. First, attitude changes in flying and particular roll movements of aircraft are dynamic events, which develop continuously across time with a more or less high angular velocity. Usually, this velocity is not very high but might increase considerably in case of system failures or even damages. Thus, providing pilots the opportunity to observe the dynamic and progressive change of the attitude on their AI and, thus, to respond quickly and early to such changes seem to be a bit more realistic than to present sudden discrete changes. The latter just simulate the somewhat artificial situation, where pilots have not monitored the PFD for a while and then suddenly notice an extreme attitude change. Second, it was expected that observing dynamically developing attitude changes would increase the subjective time pressure to respond to recover as quickly as possible before the situation becomes too extreme, especially the faster the bank angle advances toward an unsafe flight attitude. This in turn, should raise the probability to respond spontaneously with the most intuitive response (Kahneman, 2011; Wickelgren, 1977), a mechanism proposed to have contributed to at least two recent crashes where Russian pilots were required to recover from an extreme flight attitude with a Western MH AI in a stressful situation (Eidgenoessisches Departement für Umwelt, Verkehr, Energie und Kommunikation, 2002; Interstate Aviation Committee, 2008). Thus, responses to a progressive developing attitude change should be an even more sensitive test for comparing the two AI formats with respect to how intuitive they are.

Two experiments were conducted. In the first one, flight novices performed a simulated flight path tracking task, as well as different recoveries from a discrete attitude change (±45°) and continuous attitude changes with different angular velocities, respectively. The tasks had to be performed with both AI formats and displays designed in a way that closely resembled typical AIs currently in use in modern glass cockpits. Based on our previous research (Müller et al., 2018), we expected that novices would perform both tasks better with the MA compared with the MH format. Because of the assumed higher time pressure perceived with continuously developing attitude changes, we further expected that the performance differences between the two formats would become most pronounced for recoveries from continuous attitude changes, dependent on the angular velocity.

The second experiment replicated the first one but with (MH) trained pilots as participants. With respect to tracking, that is, a task which these pilots were most used to perform with “their” display from flying, we expected they would perform better (or at least equally well, see Cohen et al., 2001) with the MH than the MA format. Predictions, how the pilots would perform recoveries from discrete versus continuous attitude changes were more difficult to derive, though. On the one hand, even relatively fast continuous attitude changes might represent situations that pilots find generally more familiar than sudden discrete changes. This might make it easier for them to respond to such changes quickly and accurately with either display. On the other hand, seeing the dynamic change and the assumed raised time pressure should also cause them to respond more quickly and intuitively than they usually do. In this case, they also would be expected to perform better with the more intuitive MA display.

In both experiments not only flight performance measures were considered, but also subjective ratings of workload experienced when performing the various tasks with the different AI formats. This was done to explore whether participants subjectively perceived performing the tasks with one of the AI formats as easier and less loading than with the other. This would be particularly informative, if participants tried to compensate differences in difficulty between the two AI formats by some extra effort.

The general method of both experiments is presented hereafter. Then, the specific method, the results, and a detailed discussion for each experiment is given. To summarize the article, a discussion of both experiments with conclusion is offered.

General Method

Apparatus

A PC-based flight simulator was used to carry out the experiments. The research simulator had a wooden mock-up cockpit panel that was inspired by a Cessna 172 Skyhawk SP G1000. A PFD was presented on a screen that was implemented in the mock-up. For both AI formats the general layout of the PFD was designed to reproduce the typical design of an Airbus A320 PFD as closely as possible (see Figure 1). Only a few changes had to be made to this PFD design to meet the requirements of a use by naïve participants and the studies' task. These changes included a resign of the flight mode annunciator in the upper part of the display, the indications of control limits, and the bank indicator (aka skypointer). The latter was dismissed to provide no alternative cue than the AI for performing the recovery task. A Logitech Extreme 3D Pro joystick was used as input device and its input deflections were linearly transferred into pitch and roll rates. The simulation was a reduced linear flight model considering two degrees of freedom (pitch and roll). The outside view was generated by an X-Plane simulation and projected on the wall approximately 1.2 m behind the mock-up cockpit.

Tasks

Tracking. The participants' first task was a tracking task. Two separate random disturbance functions each axis simulated atmospheric turbulences. They were based on the sum of five sine functions (Fracker & Wickens, 1989). The participants were required to maintain a stable horizontal flight with bank and pitch angle of 0° by compensating the disturbances with inputs on *x* and *y*-axis of the joystick. The amplitude in bank was three times the amplitude in pitch.

Recovery. The second task was to recover from four different types of unexpected attitude changes: One sudden change of bank angle to ±45° (discrete stimulus) and three continuous roll movements with 45, 90, and 180°/s to the left or to the right. While presenting the stimuli, the pitch angle was at 0° but could be altered by the participants while recovering. The participant had 10 s to recover to 0° pitch and 0° bank. After finishing one trial, the next stimulus was presented after a random time interval of 5–12 s.

Design

The study designs of both experiments included a within-subject factor that represented the two reference formats of the attitude; MA and MH. A second within-subject factor contained four levels and was only used for investigating the impact of the different types of attitude changes in the recovery task (discrete change of 45°, continuous change with 45°/s, continuous change with 90°/s, and continuous change with 180°/s).

Dependent Measures

Performance in the tracking task was assessed in terms of tracking error for both axes, bank, and pitch, separately. The deflections from 0° were recorded with a frequency of 60 Hz and used to calculate the root mean square error (RMSE) in bank and pitch across trials. The lower the RMSE value, the better the participant was able to maintain a horizontal flight attitude.

Two objective performance measures were recorded to measure the performances in the recovery task, the response time and the number of reversal errors. The response time assessed the time participants needed to respond to the AI change, and was defined as the time (in milliseconds) from the first appearance of an AI change in the PFD until the first recordable movement of the flight joystick (in either direction). A reversal error was defined as an initial joystick input after the recovery stimulus that was initiated to the wrong direction; that is, an input that amplifies instead of compensates the induced bank angle change.

In addition to the performance measures, the NASA-TLX (Hart & Staveland, 1988) was used as a subjective measure to assess the perceived workload when performing the recovery tasks with the different AI formats. An overall workload score was derived by simply averaging the individual ratings across the six subscales (RAW TLX; Hart, 2006).

Data Analysis

The usual alpha-level of 5% was used for considering effects significant in the following analyses. Before data analyses, outlier corrections were made. When analyzing the tracking data, participants with a bank RMSE that was three standard deviations above the mean of the respective condition, were omitted. The tracking task's RMSE (aggregated across both 2-min parts) and the NASA-TLX data were then analyzed by one-way analysis of variances (ANOVAs) with repeated measures across the factor AI format.

Regarding objective performance measures of the recovery task, only successfully completed recovery trials were considered for the analysis. These trials included all trials where the participants were able to restabilize the attitude of the aircraft to a horizontal flight within 10 s and kept it stable for another 2 s within a range of ±2°. Additionally, for the analyses of response times, only successful trials without reversal error and a response time ≥ 100 ms were considered. To investigate the effects of the experimental factors on response time, two independent ANOVAs were used. First a 2 (AI format) × 2 (dynamics of stimulus: discrete vs. continuous) repeated measures ANOVA was calculated to evaluate the effect of a discrete versus continuous attitude changes. The data of the three angular speeds in the continuous trials were aggregated for this analysis. A second 2 (AI format) × 3 (angular speed: 45, 90, 180°/s) ANOVA was computed to evaluate specifically the impact of the angular speed of the roll movements in trials consisting of a continuous stimulus.

The analyses of reversal error data followed the same structure but based on a logistic regression approach with mixed models, taking the dichotomous nature of the reversal error data into account. A random effects variable for participants was included in the models, to control for nuisance variance of interindividual differences. Fixed factors are AI format and angular speed or dynamics of stimulus, respectively. The model selection then was performed by likelihood ratio tests, which reveals the best fitting model out of various alternative models. To calculate *p* values of the effects, likelihood ratio tests for model comparison (full model vs. model without main effect or interaction effect) were performed. The effect size ω_p^2 of each effect was calculated according to the approach of Nakagawa and Schielzeth (2010).

Experiment 1

Method

Participants. Twenty-six participants were included in the study (16 women, 10 men). None of them had any prior knowledge of flying a real aircraft, whatsoever. One participant had some limited flying experiences, based on casually flying in a flight simulator. The participants' age ranged from 18 to 65 years with a mean of 28.3 years (SD = 9.5). Every participant was compensated with either 7.00€ or course credits. The simulator study was approved by the Ethics Board of the Technische Universität Berlin's

Department of Psychology and Ergonomics and all participants were treated according to the Declaration of Helsinki.

Procedure

Each experiment began with a standardized briefing, consisting of the test procedure, the two tasks, and the flight displays. The briefing was followed by a familiarization phase (4 min) to accommodate the participants with the aircraft's simulation and controls. During the familiarization phase, the participants were required to make several flight maneuvers with turns and level flights, while only the outside view was projected as an attitude reference. This familiarization phase was included to make the flight novices familiar with the underlying logic particularly of the MH display; that is, its logic to present movements of an artificial horizon corresponding to how the changes of the natural horizon would look like when looking outside the cockpit window.

The following experiment was divided into two experimental blocks, representing the two AI format conditions. At the beginning of each block, the participants performed a training with the given AI display. This training, first, consisted of basic flying tasks to be performed with support of PFD while the outside view was still shown (about 4 min). Then, in a second phase of training (about 4 min), the outside view was disabled, and just the PFD with the AI remained as a reference to control the attitude of the aircraft.

The data collection in each block was split into two parts, each including a 2-min tracking task followed by 16 trials (4 types × 2 directions × 2 replications) of the recovery task in random order with the given AI. The block was concluded by providing NASA-TLX ratings of the perceived workload while performing the recovery task with the given AI. After a short break, the AI was changed and the second block was started with the same structure as the first one. The order of blocks (AI format) was balanced across participants. Overall, the whole session lasted about 60 min for each participant.

Results

Tracking task. One participant was considered an outlier in the tracking task. This reduces the sample size for the tracking analysis to 25 participants. No significant differences emerged in tracking performance dependent on the AI format, neither for the RMSE of bank, $F(1, 24) = 0.01$, $p = .938$, $\eta_p^2 < .01$, nor the RMSE of pitch, $F(1, 24) = 0.35$, $p = .562$, $\eta_p^2 = .01$.

Recovery task. Overall, only 3.67% of all recovery trials were excluded from data analysis because of unsuccessful recoveries or too fast responses.

Recoveries of discrete versus dynamic attitude changes. Mean response times for both AI formats in discrete versus continuous changes of the attitude are shown in Figure 2A. The participants responded significantly faster while flying with the MA display (M = 556.8 ms, SE = 18.8 ms) than with the MH display (M = 623.3 ms, SE = 19.1 ms), $F(1, 25) = 19.57$, $p < .001$, $\eta_p^2 = .44$. Additionally, they responded quicker in the continuous (M = 570.9 ms, SE = 15.8 ms) versus discrete (M = 609.1 ms, SE = 21.1 ms) condition, $F(1, 25) = 8.34$, $p = .008$, $\eta_p^2 = .25$. No significant interaction effect AI Format × Dynamics emerged, $F(1, 25) = 2.41$, $p = .133$, $\eta_p^2 = .09$.

The mean percentages of reversal error committed by the participants when performing the recoveries with the two AI formats and the different dynamics of attitude changes are shown in Figure 2B. The mixed model with an AI Format × Dynamics of Stimulus interaction of the two fixed effects and the participants as random factor with a random intercept and random slope on AI format appeared to be the best fitting model. Comparing both formats, it is apparent that participants committed overall less reversal error in the MA condition (*M* = 12.1%, *SE* = 2.6%) than in the MH condition (*M* = 24.0%, *SE* = 2.5%). This effect is reflected in a significant main effect of AI format, $\chi^2(1, N = 26) = 22.71$, $p < .001$, $\omega_p^2 = .092$. The main effect dynamics did not reach significance, $\chi^2(1, N = 26) = 0.87$, $p = .352$, $\omega_p^2 = -.002$, though. However, as expected, an interaction effect AI Format × Dynamics emerged, $\chi^2(1, N = 26) = 4.48$, $p = .034$, $\omega_p^2 = .002$. This interaction effect was reflected in participants committing considerably higher number of reversal error in response to continuous (*M* = 28.6%, *SE* = 3.3%) versus discrete changes (*M* = 19.4%, *SE* = 2.8%) when working with the MH display, while almost no difference was present for the MA display (continuous: *M* = 11.1%, *SE* = 2.5%; discrete: *M* = 13.0%, *SE* = 3.3%).

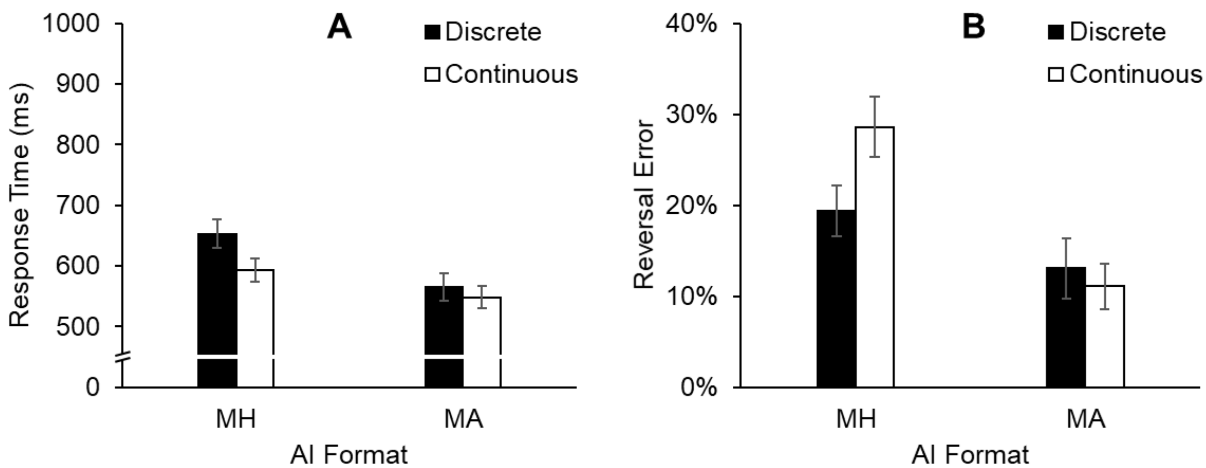

Figure 2. Novices' means of (A) response time and (B) reversal error on discrete and continuous recovery stimuli in both attitude indicator (AI) format conditions, moving horizon (MH) and moving aircraft (MA). Error bars represent standard errors.

Recoveries of continuous attitude changes with different angular speeds. Mean response times for recoveries from continuous attitude changes with different angular speeds are shown for both AI formats in Figure 3A. Independent of the speed of roll movements, participants responded generally quicker in condition MA (M = 548.3 ms, SE = 18.5 ms) compared with MH (M = 588.0 ms, SE = 17.7 ms), $F(1, 25) = 4.47$, $p = .045$, $\eta_p^2 = .15$. As expected, the response time of participants was further affected by the angular speed of the attitude change; that is, participants responded the quicker the faster the angular speed of the continuous roll movements were, $F(2, 50) = 9.38$, $p < .001$, $\eta_p^2 = .27$. No interaction effect AI Format × Angular Speed emerged, $F(1.47, 36.72) = 0.39$, $p = .617$, $\eta_p^2 = .02$.

Figure 3B depicts the mean proportion of reversal errors for the same conditions. The model including the fixed factor AI Format × Angular Speed interaction and the participants as random factor with a random intercept and random slope for AI format proved to be the best model fit. Over all continuous stimuli, significant fewer reversal error were committed in condition MA (M = 11.1%, SE = 2.5%) compared with MH (M = 28.6%, SE = 3.3%), $\chi^2(1, N = 26) = 26.38$, $p < .001$, $\omega_p^2 = .124$. In addition, and mirroring the effects in response time, significant more reversal error occurred with faster angular speed of roll movements, $\chi^2(2, N = 26) = 8.25$, $p = .016$, $\omega_p^2 = .009$. However, no significant interaction AI Format × Angular Speed emerged, $\chi^2(2, N = 26) = 0.36$, $p = .836$, $\omega_p^2 = -.002$.

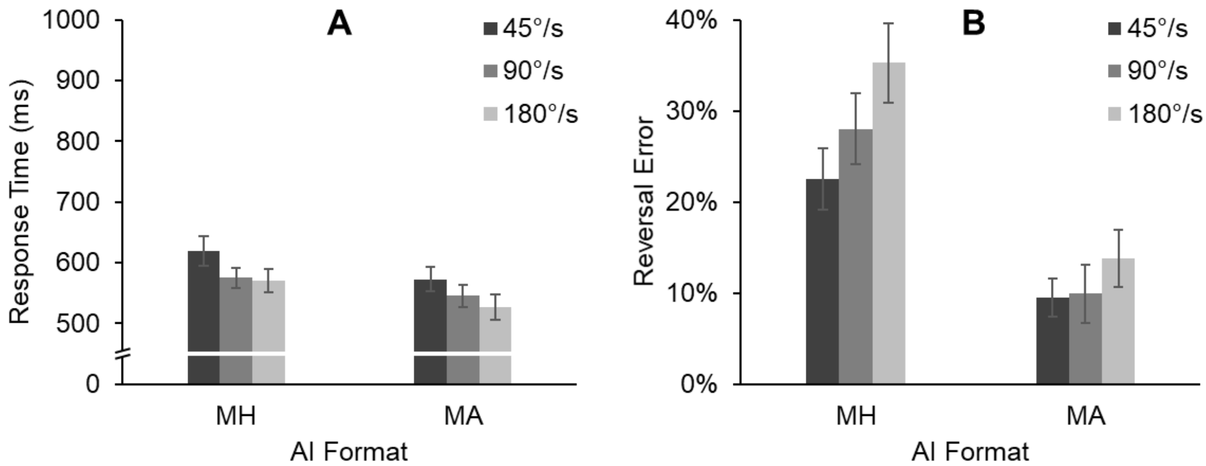

Figure 3. Novices' means of (A) response time and (B) reversal error for each angular speed of continuous recovery tasks in both attitude indicator (AI) format conditions, moving horizon (MH) and moving aircraft (MA). Error bars represent standard errors.

Workload. Because of missing data, one participant had to be excluded from this analysis of the NASA-TLX data. The mean ratings of the NASA-TLX after each AI format block did not reveal a significant effect (MH: M = 40.6, SE = 3.8; MA: M = 38.2, SE = 3.9), $F(1, 24) = 2.14$, $p = .157$, $\eta_p^2 = .08$.

Discussion

The results of the first experiment based on flight novices confirm our previous findings, pointing to an advantage of the MA versus MH AI format for this group, even when currently used PFD designs are considered (Müller et al., 2018). However, this time, the benefit of the MA format only emerged in the recovery tasks, but not in tracking. The latter "no-effect" is more in line with the findings of Yamaguchi and Proctor (2010) who also did not find any differences in tracking performance of novices supported by different AI formats. This suggests, that even for novices, differences between the two AI formats in respect to supporting tracking performance might not be large and not consistently be found.

So much the clearer differences between the two AI format emerged for the recovery tasks. Going beyond the evidence gained from previous studies (Ding & Proctor, 2017; Lee & Myung, 2013; Müller et al., 2018), the results suggest that the benefits of the MA format for this type of task are not limited to recoveries from discrete attitude changes, but becomes even more pronounced, when the unexpected roll movements develop continuously with a given angular speed. Such effect was expected, because continuously developing attitude changes should raise the perceived time pressure to respond, which, in turn, should cause a tendency to respond more intuitively to such dynamic compared with discrete attitude changes. Directly in line with this assumption is the finding of overall quicker response times for the continuous compared with the discrete attitude changes and the significant increase of response speed with higher angular speeds. Thus, the increase of reversal errors in the continuous condition dependent on the angular speed seems to reflect a basic speed–accuracy trade-off (Wickelgren, 1977). Given that higher time pressure can be expected to lead to more intuitive responding (Kahneman, 2011), this effect provides even stronger evidence than previous studies that the moving-part compatibility principle underlying the design of the MA format, indeed, is more effective than the alternate principle of pictorial realism underlying the design of the MH format in supporting natural response tendencies. This also further supports the theoretical explanations provided by Janczyk et al. (2015) and Previc and Ercoline (1999) in terms of the response-effect compatibility and the neuropsychological reference model, which both argue for the MA display to be generally superior. On a subjective level, the differences between the two AI formats were not mirrored in the overall workload rating. This suggests, that the participants were not aware that the MA format made it easier for them to respond, or have compensated for differences by any extra effort. However, because the NASA-TLX ratings were not sampled for single conditions, no clear link can be made to the different types of attitude changes included in the study.

Whether similar time-pressure effects induced by continuous attitude changes would also become visible in experienced pilots who are well trained with the MH format and presumably more familiar with continuous than discrete attitude changes was addressed in the second experiment.

Experiment 2

Method

Participants. Twelve certified airline pilots (1 woman, 11 men) participated in the study. The pilots' age ranged from 25 to 63 years (*M* = 39.3, *SD* = 13.4). Their experience ranged from 220 to 25,000 total flight hours with a mean of 8,186 hr (*SD* = 9,116). All pilots were trained to fly according to instrument flight rules (IFR). They all received their IFR training and experience with MH formatted displays or instruments. The pilots volunteered their time to participate in the study. Also this study was approved by the Technische Universität Berlin's Ethics Board and was conducted in accordance with the Declaration of Helsinki.

Procedure. The procedure of the second study corresponded largely to the procedure of the first study. However, because of higher experience and knowledge, familiarization and practice phase was considerably shortened for pilots. The briefing was condensed, and the phases to familiarize to the simulator and the AI format conditions were reduced to one session showing outside view and PFD at the same time. Overall, the experimental session lasted about 45 min for each pilot.

Results

Tracking task. Corresponding to the results for novices, the pilots did not show any differences in tracking performance, either. No significant AI format effects were observable in the analysis of RMSE for bank, $F(1, 11) = 0.62$, $p = .447$, $\eta_p^2 = .05$, and for pitch, $F(1, 11) = 0.01$, $p = .920$, $\eta_p^2 < .01$.

Recovery task. Altogether 2.08% of all recovery trials were either unsuccessful recoveries or to fast responses and, therefore, omitted from data analysis.

Recoveries of discrete versus dynamic attitude changes. Figure 4A shows the response time of the pilots in both AI format conditions in discrete and continuous recovery tasks. Considering the factor AI format, the pilots responded significantly faster in condition MA ($M = 485.2$ ms, $SE = 17.6$ ms) compared with MH ($M = 566.6$ ms, $SE = 36.0$ ms), $F(1, 11) = 10.89$, $p = .007$, $\eta_p^2 = .50$. Neither the main effect of factor dynamics, $F(1, 11) = 1.93$, $p = .193$, $\eta_p^2 = .15$, nor the interaction effect AI Format × Dynamics were significant, $F(1, 11) = 4.29$, $p = .063$, $\eta_p^2 = .28$.

Figure 4B shows the mean percentages of reversal errors committed by the pilots when performing the recoveries. Again, the model with an AI Format × Dynamics of Stimulus interaction of the fixed effects and the participants as random factor with a random intercept and random slope on AI format fitted best. Neither the main effect of format, $\chi^2(1, N = 12) = 2.37$, $p = .124$, $\omega_p^2 = .064$, nor the main effect Dynamics, $\chi^2(1, N = 12) = 0.37$, $p = .545$, $\omega_p^2 = -.007$, became significant. However, the interaction effect became significant, $\chi^2(1, N = 12) = 8.68$, $p = .003$, $\omega_p^2 = .022$. Pilots committed more reversal errors in response to the continuous stimulus ($M = 20.6\%$, $SE = 3.9\%$)

compared with the discrete stimulus (*M* = 9.0%, *SE* = 3.7%) in condition MH. Yet, they made slightly fewer reversal errors in response to the continuous stimulus (*M* = 6.7%, *SE* = 2.2%) compared with the discrete stimulus (*M* = 11.8%, *SE* = 5.5%) in condition MA.

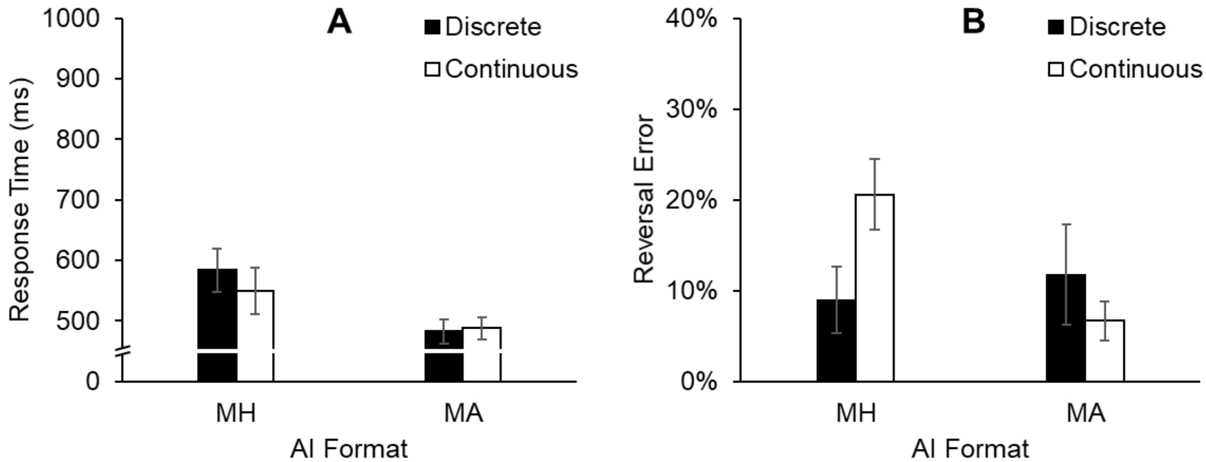

Figure 4. Pilots' means of (A) response time and (B) reversal error on discrete and continuous recovery stimuli in both attitude indicator (AI) format conditions, moving horizon (MH) and moving aircraft (MA). Error bars represent standard errors.

Recoveries of continuous attitude changes with different angular speeds. Mean response times of pilots for recoveries of continuous attitude changes with different angular speeds are shown in Figure 5A. As becomes evident, the pilots responded generally faster when working with the MA (*M* = 487.7 ms, *SE* = 18.4 ms) compared with the MH display (*M* = 550.1 ms, *SE* = 38.8 ms), reflected in a main effect of AI format, $F(1, 11) = 7.51$, $p = .019$, $\eta_p^2 = .41$. However, neither the main effect of angular speed, $F(1.62, 17.81) = 0.45$, $p = .604$, $\eta_p^2 = .04$, nor the AI Format × Angular Speed interaction became significant, $F(1.24, 13.65) = 1.89$, $p = .192$, $\eta_p^2 = .15$.

Figure 5B shows the descriptive data of the analysis on angular speeds in terms of reversal error. Again, the model including the fixed factor AI Format × Angular Speed interaction and the participants as random factor with a random intercept and random slope on AI format proved to be the best model fitting. Comparing the formats over all continuous stimuli, significantly fewer reversal error occurred in condition MA (M = 6.7%, SE = 2.2%) compared with the MH condition (M = 20.6%, SE = 3.9%), $\chi^2(1, N = 12) = 8.27$, $p = .004$, $\omega_p^2 = .143$. Additionally, significantly more reversal error occurred with faster angular speed of roll movements, $\chi^2(2, N = 12) = 8.06$, $p = .018$, $\omega_p^2 = .047$. The mean error rates depicted in Figure 5B suggests, that this latter effect was mainly because of the rise of error rates in the 90 and 180°/s conditions compared with the slowest 45°/s condition. No significant interaction effect of AI Format × Angular Speed emerged, $\chi^2(2, N = 12) = 0.29$, $p = .864$, $\omega_p^2 = .012$.

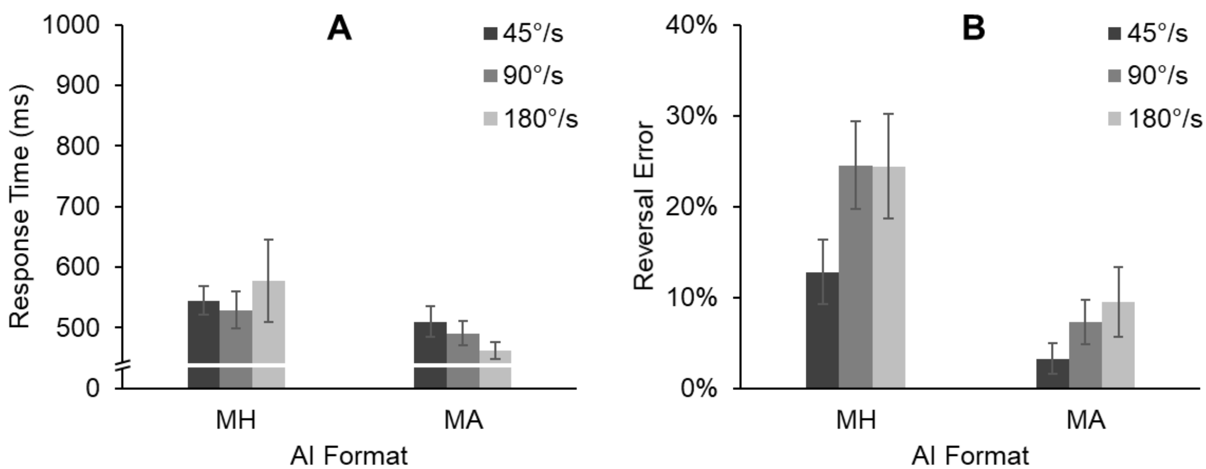

Figure 5. Pilots' means of (A) response time and (B) reversal error for each angular speed of continuous recovery tasks in both attitude indicator (AI) format conditions, moving horizon (MH) and moving aircraft (MA). Error bars represent standard errors.

Workload. The rating of the pilots' perceived workload via NASA-TLX after each AI format block did not differ significantly in the overall score (MH: M = 27.6, SE = 3.7; MA: M = 23.5, SE = 2.8), $F(1, 11) = 0.76$, $p = .401$, $\eta_p^2 = .07$.

Discussion

Corresponding to the results of our novices sample, the tracking task did not reveal significant effects between AI formats. In the case of pilots, this was expected because of their higher familiarity and practice with flight path tracking based on a MH display and the fact, that even a small benefit of the MH compared with the MA display was found for pilots performing a tracking task in our previous research (Müller et al., 2018). However, when they needed to recover quickly from unusual attitudes, even in this group, effects in favor of the MA display again became apparent. In the discrete condition, this was reflected in quicker response times with almost unchanged proneness for reversal errors. In the continuous condition, the impact of the AI format on performance was even more pronounced. Here, the pilots, although originally trained with the MH display, were able to respond quicker and also more reliably with the unfamiliar MA than with "their" MH display. However, the raised time pressure to respond to continuous attitude changes that we assumed, did not manifest in the data of pilots. In contrast to the novices (Experiment 1), the pilots responded as quickly to the continuous as to the discrete attitude changes. The fact, that this nevertheless lead to higher error rates with the MH than the MA display suggests, that looking at a continuously developing attitude change on the MH display has increased the tendency to respond intuitively with a correcting countermovement, instead of a movement in the same direction that would have been the proper response for this display. In contrast, observing the same continuous change on the MA display led to even higher portion of correct responses than in the discrete condition. These effects emerged even though the MH display used in this study was integrated in a PFD highly similar to the one they were used to from flight training and daily practice. Again, this provides strong evidence for the MA display being considerably more compatible with natural response tendencies than the MH display, if the task to be performed is relatively unusual and only know from simulator lessons. Overall, the findings of this second experiment does not only confirm the results in favor of the MA display also for pilots reported by Müller et al. (2018, Experiment 2) but also extend the conclusions to continuously evolving attitude changes. They also provide another empirical support for early observations of Kovalenko (1991) pointing to the fact, that even pilots trained with the MH display obviously need increased cognitive effort to interpret it correctly compared with the MA display.

Remarkably, also for the pilot sample, the performance difference in the recovery task were not reflected in the subjective workload ratings. This suggests that also the pilots were subjectively not aware of the higher cognitive demand of the MH format. This might present a risk because they obviously do not realize their susceptibility to errors with this display. A similar result was found by Müller et al. (2019) in a study with highly trained flight novices when transferring from MH to MA format and vice versa.

Summary and Concluding Discussion

The present research revisits the classic human factors issue of what reference format of the AI in commercial aircrafts might be most useful to support the spatial orientation of pilots in case no outside view is available. In contrast to many earlier studies, this research included a more realistic set of nominal and off-nominal flight tasks, which had to be performed relying on actual AIs of modern glass cockpit PFDs.

Taken together, both experiments provided another evidence for the superiority of the MA format over the MH format, at least for flight tasks, which are somewhat unusual and/or require quick and strong responses. Therefore, while no differences between the two formats were found when novices or pilots were requested to perform a simulated flight path tracking with only small deviations, which need to be corrected, a clear benefit of the MA display emerged for recoveries from unusual attitudes. Thus, these findings add further support for the theory of response-effect compatibility (Janczyk et al., 2015) and neuropsychological reference model (Previc & Ercoline, 1999).

In contrast to previous research, the current study did not only investigate recovery performance in response to sudden discrete attitude changes but also to continuously evolving ones, what pilots might find more familiar from real flight situations. For both groups, these dynamically evolving attitude changes made the differences between both AI formats even clearer. As expected, novices tended to respond faster to these changes, which led to higher error rates in their initial movements. For pilots no evidence of raised time pressure was observed in response times, but they nevertheless became more error prone when recovering from continuous changes. This is in line with our assumptions that dynamic attitude changes provoke a more intuitive responding, which then makes the MA format advantageous, but it is not entirely certain whether this can be attributed to a greater perceived time pressure. Overall, these results suggest that the advantage of the MA compared with the MH display indeed is superior, because it better fits to natural response tendencies (cf. also Previc & Ercoline, 1999). Confirming similar results of Müller et al. (2018), this shows that early results of the 1950–1970s, pointing in the same direction, were not just specific for the relatively small instruments of that time but also apply to the larger PFD design of current aircraft.

In the pilot sample, the benefits of the MA compared with the MH display emerged although the two AIs were implemented in typical PFDs used in current glass cockpits and even though the pilots were highly trained with the MH format. This suggests, that even MH trained pilots might have no difficulties if they transferred from "their" MH to the MA format; that is, transferred from an incompatible to a more compatible and intuitive display format. However, the reverse transfer from MA to the MH format can involve much bigger issues (Kovalenko, 1991; Müller et al., 2019) and even lead to severe consequences for flight safety, as indicated by recent flight accidents (Interstate Aviation Committee, 2008; Eidgenoessisches Departement für Umwelt, Verkehr, Energie und Kommunikation, 2002). Thus, the results can be taken as empirical evidence that a general change of the AI format from the current MH standard to the MA format in future aircraft might be beneficial and possible without much risk. In any case, the MA format should be seriously considered as the standard AI with regard to ground control stations of unmanned aerial vehicles.

However, as this research also demonstrates, recommendations concerning specific displays should not only be based on a limited range of task and limited set of applications. Instead, all relevant flight tasks and contexts, which might be affected by the change, must be examined together as comprehensively as possible. The current recommendation in favor of the MA format has been based on investigating three different tasks, including a simulation of flight path tracking and recoveries from unexpected continuous and discrete attitude changes. All of these tasks can be considered as representing typical demands pilots have to cope with in nominal or off-nominal situations. Together with the fact that we used AI displays, which were as similar as possible to typical AIs in modern PFDs, we assume that the results are generalizable to real flying.

However, the tasks used certainly do not include all typical tasks to be performed with support of the AI. For a complete understanding of performance consequences, future research should also consider other flight tasks, for example, the active adjustment to a given attitude change (see for a first example Janczyk et al., 2015) or dynamic transitions from ordinary flight path tracking to a recover situation. In addition, before generally accepting the advantages of the MA format, it will be necessary to investigate its effects as part of head-up displays or synthetic vision displays. These displays present the AI superimposed on the natural horizon or at least a synthetic depiction of it, which might involve new compatibility issues between the movements of the AI and the movements of the natural horizon line.

References

Beringer, D. B., & Ball, J. D. (2009). Unknown-attitude recoveries using conventional and terrain-depicting attitude indicators: Difference testing, equivalence testing, and equivalent level of safety. *The International Journal of Aviation Psychology, 19,* 76–97. http://dx.doi.org/10.1080/10508410802597366

Beringer, D. B., Williges, R. C., & Roscoe, S. N. (1975). The transition of experienced pilots to a frequency-separated aircraft attitude display. *Human Factors, 17,* 401–414. http://dx.doi.org/10.1177/001872087501700411

Browne, R. C. (1954). Figure and ground in a two dimensional display. *Journal of Applied Psychology, 38,* 462–467. http://dx.doi.org/10.1037/h0057045

Cohen, D., Otakeno, S., Previc, F. H., & Ercoline, W. R. (2001). Effect of “inside-out” and “outside-in” attitude displays on off-axis tracking in pilots and nonpilots. *Aviation, Space, and Environmental Medicine, 72,* 170–176.

Ding, D., & Proctor, R. W. (2017). Interactions between the design factors of airplane artificial horizon displays. *Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting, 61,* 84–88. http://dx.doi.org/10.1177/1541931213601487

Dunlap & Associates, Inc. (1955). *Pilot performance with two different attitude displays* (Rep. No. DTIC No. AD-082 868). Washington, DC: Office of Naval Research.

Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation. (2002). *Schlussbericht des Büros für Flugunfalluntersuchungen über den Unfall des Flugzeuges Saab 340B, HB-AKK, betrieben durch Crossair unter Flugnummer CRX 498* [Final report of the aircraft accident investigation bureau on the accident to the SAAB 340B aircraft, registration HB-AKK of Crossair flight CRX498]. Bern, Switzerland. Retrieved from https://www.sust.admin.ch/in-halte/AV-berichte/1781_d.pdf

Fracker, M. L., & Wickens, C. D. (1989). Resources, confusions, and compatibility in dual-axis tracking: Displays, controls, and dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 80–96. http://dx.doi.org/10.1037/0096-1523.15.1.80

Gross, A., & Manzey, D. (2014). Enhancing spatial orientation in novice pilots: Comparing different attitude indicators using synthetic vision systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58,* 1033–1037. http://dx.doi.org/10.1177/1541931214581216

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50,* 904–908. http://dx.doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam, Netherlands: North-Holland Press. http://dx.doi.org/10.1016/S0166-4115(08)62386-9

Interstate Aviation Committee. (2008). *Aeroflot Flight 821 accident report*. Moscow, Russia. Retrieved from https://www.icao.int/safety/airnavigation/AIG/Documents/Safety%20Recommendations%20to%20ICAO/Final%20Reports/ MAKA TR72VP-BYZ_final_report.pdf

Janczyk, M., Pfister, R., Crognale, M. A., & Kunde, W. (2012). Effective rotations: Action effects determine the interplay of mental and manual rotations. *Journal of Experimental Psychology: General, 141,* 489–501. http://dx.doi.org/10.1037/a0026997

Janczyk, M., Yamaguchi, M., Proctor, R. W., & Pfister, R. (2015). Response-effect compatibility with complex actions: The case of wheel rotations. *Attention, Perception, & Psychophysics, 77,* 930–940. http://dx.doi.org/10.3758/s13414-014-0828-7

Johnson, S. L., & Roscoe, S. N. (1972). What moves, the airplane or the world? *Human Factors, 14*, 107–129. http://dx.doi.org/10.1177/001872087201400201

Kahneman, D. (2011). Thinking, fast and slow. New York, NY: Farrar, Straus & Giroux.

Kovalenko, P. A. (1991). Psychological aspects of pilot spatial orientation. *ICAO Journal, 46*, 18–23.

Lee, B. G., & Myung, R. (2013). Attitude indicator design and reference frame effects on unusual attitude recoveries. *The International Journal of Aviation Psychology, 23*, 63–90. http://dx.doi.org/10.1080/10508414.2013.746536

Malcolm, R. (1983). The Malcolm horizon: History and future. In R. S. Kellogg Chair (Eds), *Proceedings of the Peripheral Vision Horizon Display (PVHD) conference* (pp. 11–40). Edwards, CA: NASA.

Müller, S., Roche, F., & Manzey, D. (2019). Attitude indicator format: How difficult is the transition between different reference systems? *Aviation Psychology and Applied Human Factors, 9*, 95–105. http://dx .doi.org/10.1027/2192-0923/a000168

Müller, S., Sadovitch, V., & Manzey, D. (2018). Attitude indicator design in primary flight display: Revisiting an old issue with current technology. *The International Journal of Aerospace Psychology, 28,* 46–61. http://dx.doi.org/10.1080/24721840.2018.1486714

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society, 85,* 935–956. http://dx.doi.org/10.1111/j.1469-185X.2010.00141.x

Previc, F. H., & Ercoline, W. R. (1999). The "outside-in" attitude display concept revisited. *The International Journal of Aviation Psychology, 9,* 377–401. http://dx.doi.org/10.1207/s15327108ijap0904_5

Previc, F. H., & Ercoline, W. R. (Eds.). (2004). *Spatial disorientation in aviation*. Reston: American Institute of Astronautics and Aeronautics. http://dx.doi.org/10.2514/4.866708

Roscoe, S. N. (1968). Airborne displays for flight and navigation. *Human Factors, 10,* 321–332. http://dx.doi.org/10.1177/001872086801000402

Roscoe, S. N., & Williges, R. C. (1975). Motion relationships in aircraft attitude and guidance displays: A flight experiment. *Human Factors, 17,* 374–387. http://dx.doi.org/10.1177/001872087501700409

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85. http://dx.doi.org/10 .1016/0001-6918(77)90012-9

Wickens, C. D. (2003). Aviation displays. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 147–200). Mahwah, NJ: Erlbaum.

Yamaguchi, M., & Proctor, R. W. (2010). Compatibility of motion information in two aircraft attitude displays for a tracking task. *The American Journal of Psychology, 123,* 81–92. http://dx.doi.org/10.5406/amerjpsyc .123.1.0081